Supplementary Material to:

Recurrent fusions in *MYB* and *MYBL1* Define a Common, Transcription Factor-Driven Oncogenic Pathway in Salivary Gland Adenoid Cystic Carcinoma

Kathryn J. Brayer, Candace A. Frerich, Huining Kang and Scott A. Ness*

Department of Internal Medicine

University of New Mexico Health Sciences Center

Albuquerque, NM 87131-0001 USA

*** Corresponding Author:**

Scott A. Ness

Department of Internal Medicine, Division of Molecular Medicine

UNM Health Sciences Center and UNM Cancer Center, MSC07 4025 – CRF 121

1 University of New Mexico, Albuquerque, NM 87131-0001 USA

Tel. (505) 272-9883

Email: sness@salud.unm.edu

FIGURE S1

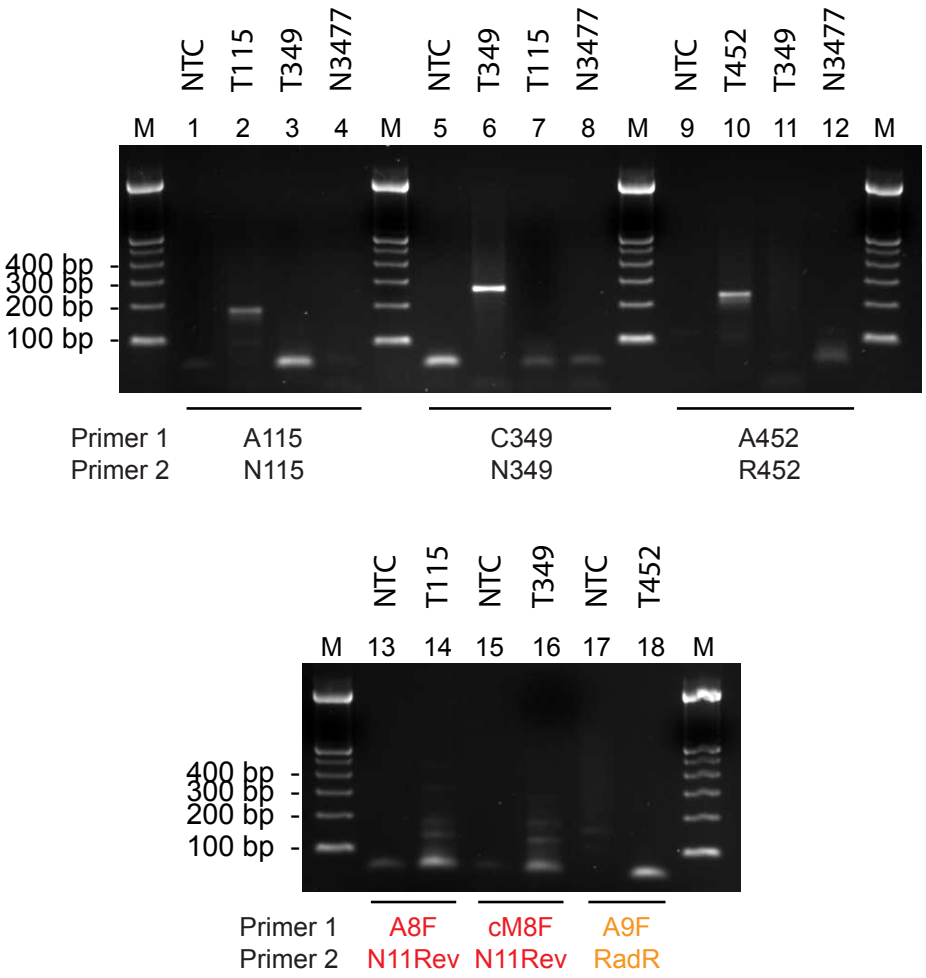

Figure S1. Verification of MYBL1 Translocation in Genomic DNA

Genomic DNA was extracted from new FFPE slides for patient samples T115, T349 and T452 and normal salivary gland sample N3477. PCR amplifications were performed using primer pairs targeting the chromosome breakpoints detected by RNA-seq for each sample, as described in the Supplementary Methods section, then products were separated on a 1.5% agarose gel. Primers A115 and N115 amplified a MYBL1-NFIB fusion product from tumor T115 (lane 2) but not the no-template control (NTC) or the other samples. Primers C349 and N349

amplified a MYB-NFIB fusion product from genomic DNA from sample T349 (lane 6) but not the controls or the other samples, and primers A452 and R452 amplified a MYB-RAD51B fusion product only from the genomic DNA from sample T452 (lane 10). The bands from lanes 2, 6, and 10 were Topo TA cloned and subjected to conventional (Sanger) sequencing for verification (see Supplementary Table S3). The reactions in the lower panel (lanes 13-18) wre performed with the same genomic DNA samples but using the primers designed for RT-PCR (see Figures 2B and 2C). All were negative, ruling out the possibility of RNA contamination.

FIGURE S2

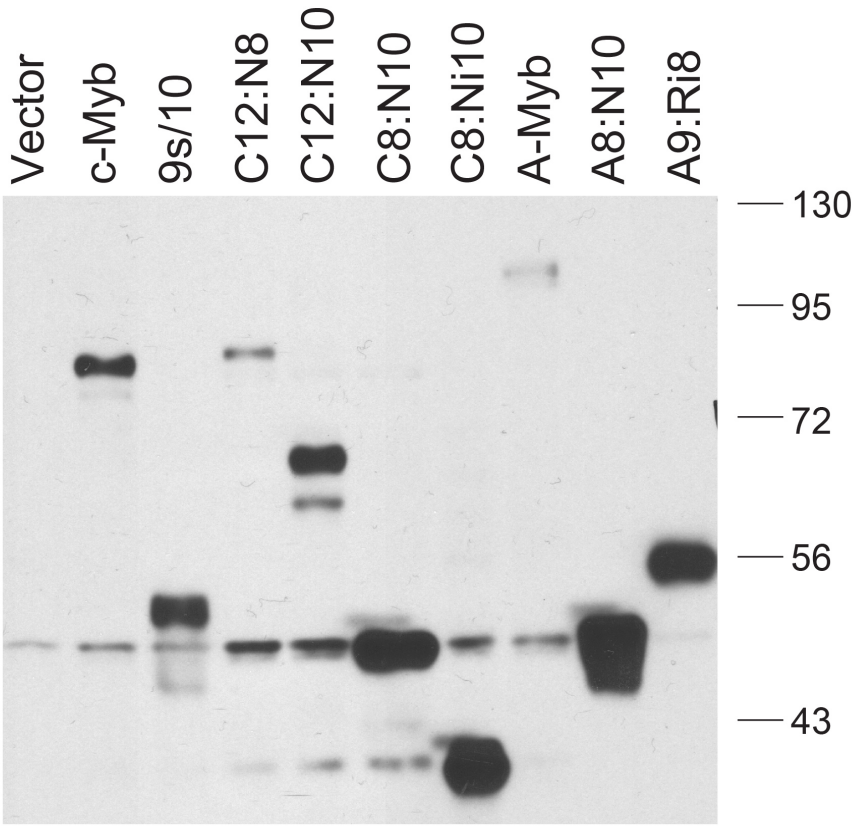

Figure S2. Expression of Recombinant c-Myb and A-Myb Fusion Proteins.

Western blot analysis of HEK293T cells transfected with empty vector (pcDNA 3.0), or plasmids expressing c-Myb, the c-Myb splice variant 9S/10, A-Myb, or fusion proteins detected in ACC tumors and described in the text and in Figure 3. Cells were transfected and total protein was harvested as described in the Methods. Cell extracts were fractionated by 10% SDS-PAGE electrophoresis and Myb proteins were transferred to a membrane and detected using rabbit

antiserum specific for the c-Myb DNA binding domain (1). Numbers along the right side indicate the molecular weights (kD) and migration of size markers.

FIGURE S3

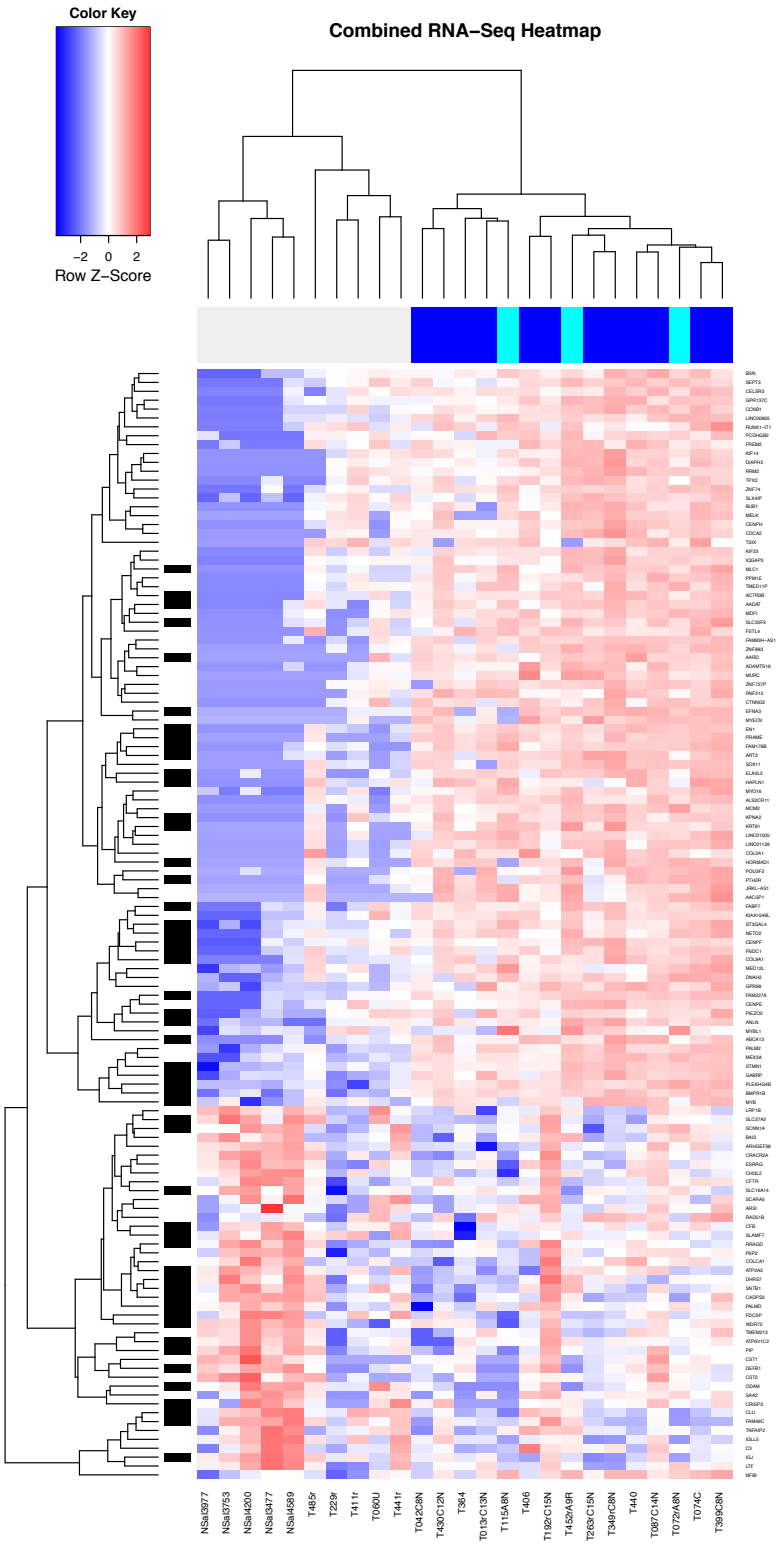
Color Key
−2 0 2
Row Z−Score
Combined RNA−Seq Heatmap
NSal3977
NSal3753
NSal4200
NSal3477
NSal4589
T485r
T229r
T411r
T060J
T441r
T042C8N
T430C12N
T364
T013rC13N
T115A8N
T406
T192rC15N
T452rA9R
T265rC15N
T349rC8N
T440
T087C14N
T072rA8N
T074C
T395C8N

Figure S3. Gene Expression Analysis

Heatmap summarizing differentially expressed genes in ACC tumors vs. normal salivary gland samples. Shading indicates range of fold-change differences in up (red) or down (blue) expression. Sample names are shown at the bottom, and the color bar along the top indicates samples with either *MYB* (dark blue) or *MYBL1* (cyan) translocations. Black bars at the left indicate genes that were reported by Gao et al. (2) to be differentially expressed in ACC tumors using high quality RNA and microarray assays. Labels along the right side show gene names (gene symbols) for the selected, differentially expressed genes. The complete sets of differentially expressed genes are detailed in Supplementary Tables S5 and S6.

FIGURE S4

NAME	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX	LEADING EDGE
GAO_TOP200	109	0.44541	1.56783	0.04704	0.15490	0.115	488	tags=46%, list=31%, signal=62%
GAO_TOP500	230	0.43391	1.56673	0.01765	0.07745	0.115	369	tags=36%, list=23%, signal=40%

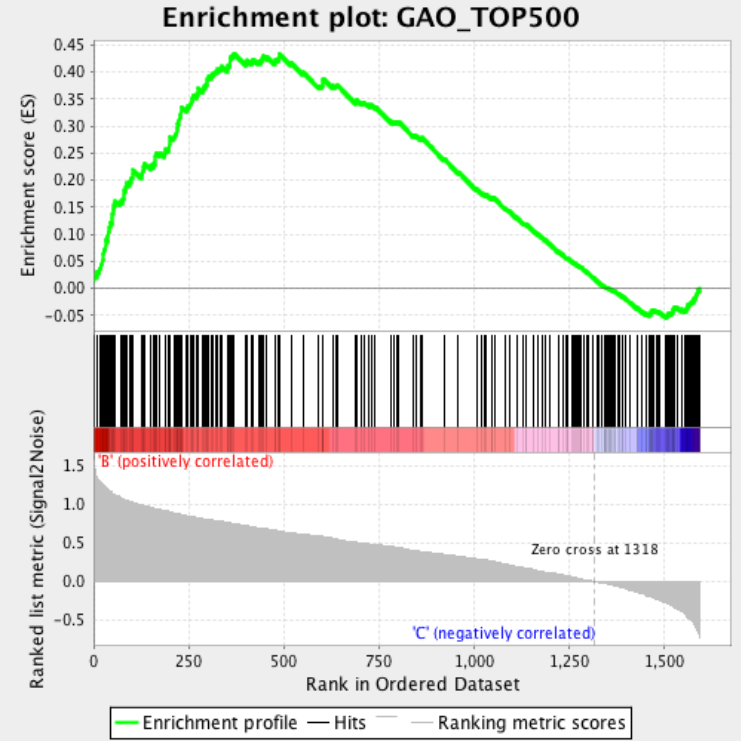

Figure S4. Gene Set Enrichment Analysis

Differentially expressed genes identified by RNA-seq (1596 genes: Group 1 vs. Normal) were compared to the top 200 or top 500 genes identified in ACC tumors by microarray analysis (2) using the Gene Set Enrichment Analysis (GSEA) tool (3). TOP: The table shows the statistics generated by the GSEA analysis. Both lists passed the recommended maximum of FDR q-val < 0.25. BOTTOM: Enrichment plot showing Enrichment Score plotted vs. Rank in Ordered Dataset. For details of the GSEA analysis see (3)

FIGURE S5

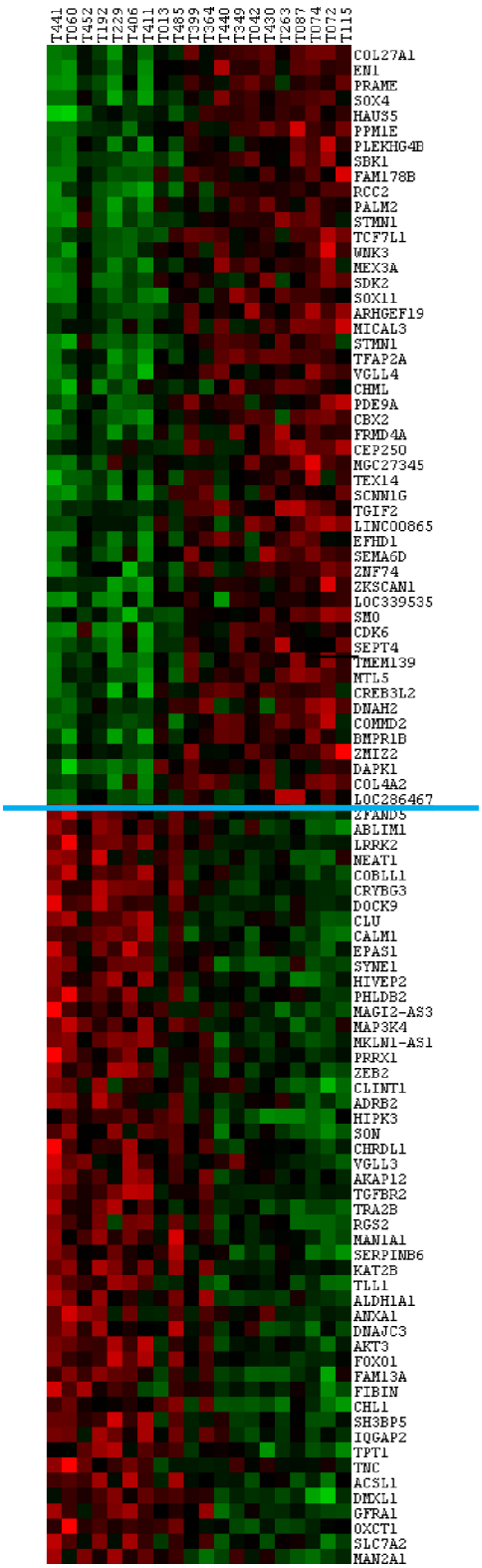
T441
T060
T452
T192
T229
T406
T411
T013
T485
T399
T364
T440
T349
T042
T430
T263
T087
T074
T072
T115
COL27A1
EN1
PRAME
SOX4
HAUS5
PPM1E
PLEKHG4B
SBK1
FAM178B
RCC2
PALM2
STMN1
TCF7L1
WNK3
MEX3A
SDK2
SOX11
ARHGEF19
MICAL3
STMN1
TFAP2A
VGLL4
CHML
PDE9A
CBX2
FRMD4A
CEP250
MGC27345
TEX14
SCNN1G
TGIF2
LINC00865
EFHD1
SEMA6D
ZNF74
ZKSCAN1
LOC339535
SMO
CDK6
SEPT4
TMEM139
MTL5
CREB3L2
DNAH2
COMMD2
BMPR1B
ZMIZ2
DAPK1
COL4A2
LOC286467
ZFAND5
ABLIM1
LRRK2
MEAT1
COBLL1
CRYBG3
DOCK9
CLU
CALM1
EPAS1
SYNE1
HIVEP2
PHLDB2
MAGI2-AS3
MAP3K4
MKLN1-AS1
PRRX1
ZEB2
CLINT1
ADRB2
HIPK3
SON
CHRDL1
VGLL3
AKAP12
TGFBR2
TRA2B
RGS2
MAN1A1
SERPINB6
KAT2B
TLL1
ALDH1A1
ANXA1
DNAJC3
AKT3
FOXO1
FAM13A
FIBIN
CHL1
SH3BP5
IQGAP2
TPT1
TNC
ACSL1
DMXL1
GFRA1
OXCT1
SLC7A2
MAN2A1

Figure S5. Heatmap of Genes Correlated to Combined *MYB* and *MYBL1* Levels

Heatmap of expression levels of the top 50 genes positively (above the line) and negatively (below the line) correlated with combined *MYB* and *MYBL1* expression. Samples are ranked by total *MYB* + *MYBL1* expression from left to right (low to high) with sample names indicated along the top of the heatmap. Color indicates whether genes are up- (red) or down-regulated (green). See Supplementary Methods for analysis details.

SUPPLEMENTARY MATERIALS AND METHODS

RNA Isolation

Formaldehyde Fixed Paraffin Embedded (FFPE) samples of ACC tumors or normal salivary gland were provided as 10 micron sections baked onto "+" slides by the tumor repositories. A 0.5-1.0 cm^2 area of paraffin-embedded tissue was scraped from a single slide into an RNase-free microcentrifuge tube for processing. Total RNA was isolated for each salivary gland sample (both tumor and normal) using Qiagen's RNAeasy FFPE kit with minor modifications. Paraffin was removed by adding 1 mL of CitriSolv, mixed on a vortex mixer for 10 seconds, then centrifuged at full speed for 2 minutes. CitriSolv was removed from the pellet, which was then washed with 1 mL 100% ethanol, and finally mixed on a vortex mixer and centrifuged as described above. Residual ethanol was removed by drying the pellet at 37 °C for ~10 minutes. The remaining steps in the manufacturer's protocol were followed exactly, and RNA was eluted from the column in 22 μL RNAse-free water. The quality of the isolated RNA was determined using the Agilent RNA 6000 Pico chip on the Agilent Bioanalyzer, while concentration was determined either by the Agilent RNA 6000 Pico chip on the Agilent Bioanalyzer or by Qubit 2.0 Fluorometer with the RNA Assay kit (Thermo Fisher).

RNA-seq Library Preparation

After RNA isolation, ~100 ng total RNA was mixed with 2 μL of 1:1000 dilution ERCC spike in control 1 (Life Technologies) and converted to cDNA using the SMARTer Universal Low Input RNA kit (Clontech) following the manufacturer's protocol with minor modification. Briefly, 50 or 100 ng RNA was mixed with either 1 or 2 μL (respectively) of a 1:1000 dilution of ERCC spike in

control 1. When necessary, volume was brought up to 10 μL with RNAase-free water. RNA-ERCC mixture was then converted to cDNA using random primers provided with the kit, at 42 °C for 90 minutes. After the suggested cleanup using AMPure beads (Agencourt), cDNA was amplified following the suggested protocol except increasing the number of cycles from 10 to 15. After cleanup, the entire cDNA reaction was barcoded and amplified following the "Life Technologies Demonstrated Protocol: Ion ChIP-Seq Library Preparation", beginning at the "Ligate adapters, nick repair, and purify the ligated ChiP DNA" step. All libraries were checked for size distribution and quality on an Agilent High Sensitivity DNA chip (Agilent Technologies). Libraries were quantified by either qPCR using the Ion Library TaqMan Quantitation kit (Thermo Fisher), or using an Agilent High Sensitivity DNA chip. Only libraries of appropriate size distribution (minimum of 100 bp fragments, avg. 200 bp fragments) were diluted to 100 pM and four libraries were pooled in equimolar concentration prior to sequencing by the Analytical and Translational Genomics Shared Resource (University of New Mexico Cancer Center) on an Ion Proton Sequencer using a P1v2 chip (Thermo Fisher).

Analysis Methods

RNA-seq reads were aligned to the human genome (GRCh37; hg19) using two separate aligners, TMAP (v4.0.6) and STAR (v 2.3.0e_r291); in both cases alignments were mapped to a BED file containing non-overlapping exons from UCSC genome hg19 (see TableS1). *MYB* and *MYBL1* translocations were visualized by importing .bam files into the Integrative Genomics Viewer (v2.3.52) (4). All further analysis was done using aligned data, and exon counts were generated using HT-Seq (5) with the FeatureCounter plugin on the Ion Torrent Server (v1.0.6, Mode = union) against the exon BED file described above.

Gene counts were generated by summing counts across exons. Counts were normalized for technical variation to ERCC standard curves determined using the ERCC Analysis plugin (v4.2-r87667). Specifically, the slope and y-intercepts generated by the ERCC report were used to recalculate counts using the formula: [new count = slope x old count + y-intercept]. ERCC slopes and y-intercepts for each sample are reported in Table S1. Samples were also normalized for library size as calculated by edgeR (values reported in Table S1). Total *MYB* and *MYBL1* expression for specific samples were calculated based on counts to DBD coding exons (exons 4-6), which had been normalized to that sample's GAPDH expression (exons 5, 7, 8). Based on total *MYB* and *MYBL1* expression, samples were sorted into two groups: high (Group 1) or low (Group 2) expressing. Finally, a minimum expression threshold of 20 reads in at least 3 samples was used to exclude low-expressing genes. Principal component analysis was done using edgeR and DESeq in R/Bioconductor (6-10). Differentially expressed genes were obtained by crosswise comparison of three groups: Normal salivary, G1 tumors, and G2 tumors using the glm method in edgeR with an adjusted p-value cutoff of 0.05 and requiring a minimum of 2-fold change. SAM (samr package; (11, 12)) was used to identify the genes whose expression were correlated with *MYB* and *MYBL1*. The significance level false discovery rate was (FDR) = 0.05 after first normalizing and variance stabilizing the data. PANTHER (release 20150430) was used to obtain GO annotations for co-expressed genes (13, 14).

RT-PCR

Total RNA extracted as described above was converted to cDNA and then amplified using the SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase using the

standard protocol. Each reaction contained ~3 ng of total RNA (extracted as described above) and gene specific primer pairs (Table SX) ordered from Integrated DNA Technologies. Primer pair annealing temperatures were based on primer Tm's and were either 46 °C (*MYB* exon 12 and 15 primer pair) or 48 °C (all *MYB*/*MYBL1* and *NFIB*/*Rad51b* pairs). PCR conditions were as follows: 55 °C for 30 minutes; .94 °C for 2 minutes; followed by 40 cycles of: 94 °C for 15 seconds, 46 °C or 48 °C for 30 seconds, 68 °C for 90 seconds; then 68 °C for 5 minutes.

DNA Verification of Translocations

DNA was isolated from hematoxylin and eosin stained Formaldehyde Fixed Paraffin Embedded (FFPE) sections of ACC tumors or normal salivary gland provided as 10 micron sections baked onto "+" slides by the tumor repositories using Qiagen's QIAamp DNA FFPE tissue kit with minor modifications. Briefly, tissue was scraped from the slide into a microfuge tube containing 1 mL CitriSolv. Samples were mixed on a vortex mixer for 10 seconds, then centrifuged at full speed for 2 minutes. CitriSolv was removed from the pellet, which was then washed with 1 mL 100% ethanol, and again mixed on a vortex mixer and centrifuged as described above. Residual ethanol was removed by drying the pellet at 37 °C for ~10 minutes. Pellets were resuspended in 180 µL of ATL buffer (Qiagen) and proteinase K, mixed on a vortex mixer, and incubated overnight at 56 °C. After 1 hour of incubation at 90 °C to reverse crosslinking, RNase A was added and samples were incubated overnight at 37 °C. The remaining steps in the manufacturer's protocol were followed exactly, and DNA was eluted from column in 100 µL water. Isolated genomic DNA quality was determined using the Agilent High Sensitivity DNA kit chip on the Agilent Bioanalyzer, while concentration was determined by Qubit 2.0 Fluorometer with the ds DNA HS Assay kit (Life Technologies).

DNA translocations were verified by PCR using genomic DNA extracted as described above and sample specific primer pairs (see Supplementary Table S3). Each 20 µL reaction contained 0.4 µM primer, 0.5 mM dNTPs, 4 U native taq (Invitrogen), 2 mM MgCl2) and ~15 ng genomic DNA (15). Cycling conditions were: 95 °C for 15 min, then 35 cycles of 95 °C for 30 seconds, 58 °C for 80 seconds, 72 °C for 60 seconds, followed by 72 °C for 10 minutes. PCR fragments were then cloned (Topo TA for sequencing, Life Technologies) and insert-positive plasmids were analyzed by Sanger sequencing (Eurofins) to confirm translocations.

Luciferase Assays and Western Blot

HEK293T cells were seeded in 24 well plates with approximately 4-6x10^4 cells per well and transfected 24 hours later with 5 ng of pcDNA3.0 containing *MYB* or *MYBL1* fusion proteins (western blot and luciferase assays) and 50 ng of luciferase reporter plasmid (Promega, luciferase assays only). Transfections were performed in duplicate using the TransIT-2020 transfection (Mirus) reagent according to manufacturer instructions. For the western blot, total protein was harvested 48 hours later by boiling in SDS-PAGE loading buffer. Cell extracts were fractionated by SDS-PAGE (10%) electrophoresis and Myb proteins were transferred to a membrane and detected using rabbit antiserum specific for the c-Myb DNA binding domain (1). For the luciferase assays, cells were harvested and firefly luciferase activity was measured after 48 hours using the Luciferase Assay System (Promega). Background subtracted data was normalized to cells transfected with empty pcDNA3.0 and the reporter plasmid. Reporter gene assays were performed in triplicate.

Table S7. Gene Ontology Analysis

The up-regulated genes shown in the heatmap (Figure S3) were compared to the Gene Ontology (GO) annotations from the GO database using the PANTHER on-line tool (4). The table shows the GO Biological Process groups that were significantly enriched in the genes up-regulated in ACC tumors, as detected by RNA-seq. For details about the GO analysis see (5,6).

GO biological process complete	Set #	Found #	Expected #	Fold Enrichment	+/-	P value
Unclassified	5728	10	18.91	0.53	-	0.00E+00
cell cycle process	1001	18	3.31	> 5	+	2.33E-05
cell cycle	1280	20	4.23	4.73	+	2.80E-05
mitotic cell cycle	777	16	2.57	> 5	+	3.11E-05
mitotic cell cycle process	706	15	2.33	> 5	+	6.77E-05
nuclear division	433	11	1.43	> 5	+	1.40E-03
chromosome segregation	198	8	0.65	> 5	+	2.47E-03
organelle fission	460	11	1.52	> 5	+	2.55E-03
regulation of cell cycle process	483	11	1.59	> 5	+	4.10E-03
mitotic nuclear division	327	9	1.08	> 5	+	1.04E-02
cell division	467	10	1.54	> 5	+	2.42E-02
kinetochore assembly	10	3	0.03	> 5	+	4.20E-02

Table S8. GO analysis of genes whose expression is either positively (Top) or negatively (Bottom) correlated with combined MYB and MYBL1 expression. Only processes with >= 2.5 fold enrichment are shown. See supplemental methods for details of analysis.

GO Biological Process (complete)	No. of Genes in Genome Annotated to Process	No. of Genes on List Annotated to Process	Expected No. of Genes	Fold Enrichment	P-value
Genes whose expression is positively correlated with combined MYB expression (N=94)					
Columnar/cuboidal epithelial cell differentiation	113	7	0.51	> 5	7.16E-03
Neuron differentiation	1005	20	4.54	4.41	1.52E-04
Generation of neurons	1361	23	6.15	3.74	2.22E-04
Neurogenesis	1436	23	6.49	3.55	5.88E-04
Nervous system development	2060	25	9.3	2.69	2.54E-02
Transcription, DNA-templated	2537	30	11.46	2.62	3.16E-03
Nucleic acid-templated transcription	2538	30	11.46	2.62	3.19E-03
RNA biosynthetic process	2657	30	12	2.5	8.48E-03
Genes whose expression is negatively correlated with combined MYB expression (N=221)					
Regulation of cardiac muscle hypertrophy	28	6	0.3	> 5	5.32E-03
Regulation of muscle hypertrophy	29	6	0.31	> 5	6.51E-03
Negative regulation of MAPK cascade	138	10	1.47	> 5	2.19E-02
Regulation of G-protein coupled receptor protein signaling pathway	146	10	1.55	> 5	3.57E-02
Negative regulation of kinase activity	218	12	2.31	> 5	3.69E-02
Negative regulation of intracellular signal transduction	391	19	4.15	4.58	4.21E-04
Negative regulation of protein phosphorylation	334	16	3.55	4.51	5.87E-03
Negative regulation of phosphorylation	361	17	3.83	4.44	3.32E-03
Regulation of protein serine/threonine kinase activity	432	20	4.59	4.36	4.04E-04
Regulation of protein kinase activity	682	26	7.24	3.59	1.84E-04
Negative regulation of phosphorus metabolic process	481	18	5.11	3.52	3.77E-02
Negative regulation of phosphate metabolic process	481	18	5.11	3.52	3.77E-02
Cell activation	646	24	6.86	3.5	1.05E-03
Regulation of kinase activity	732	27	7.77	3.47	1.90E-04
Regulation of transferase activity	864	29	9.17	3.16	4.15E-04
Negative regulation of apoptotic process	829	27	8.8	3.07	2.23E-03
Negative regulation of programmed cell death	838	27	8.9	3.03	2.75E-03
Regulation of response to stress	1265	39	13.43	2.9	1.63E-05
Response to oxygen-containing compound	1250	38	13.27	2.86	3.96E-05
Negative regulation of cell death	891	27	9.46	2.85	8.84E-03
Negative regulation of protein metabolic process	913	27	9.69	2.79	1.39E-02
Negative regulation of cellular protein metabolic process	852	25	9.05	2.76	3.95E-02
Regulation of apoptotic process	1395	40	14.81	2.7	7.42E-05
Regulation of programmed cell death	1412	40	14.99	2.67	1.03E-04
Negative regulation of signal transduction	1025	29	10.88	2.66	1.34E-02
Negative regulation of molecular function	1007	28	10.69	2.62	2.84E-02
Negative regulation of response to stimulus	1271	35	13.5	2.59	1.87E-03
Regulation of cell death	1484	40	15.76	2.54	3.96E-04
Negative regulation of signaling	1121	30	11.9	2.52	2.64E-02
Regulation of protein phosphorylation	1161	31	12.33	2.51	1.87E-02
Positive regulation of catalytic activity	1355	36	14.39	2.5	2.89E-03
Negative regulation of cell communication	1130	30	12	2.5	3.09E-02

REFERENCES

1. Dash AB, Orrico FC, Ness SA. The EVES motif mediates both intermolecular and intramolecular regulation of c-Myb. Genes Dev. 1996;**10**: 1858-69.

2. Gao R, Cao C, Zhang M, Lopez MC, Yan Y, Chen Z, et al. A unifying gene signature for adenoid cystic cancer identifies parallel MYB-dependent and MYB-independent therapeutic targets. Oncotarget. 2014;**5**: 12528-42.

3. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;**102**: 15545-50.

4. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;**14**: 178-92.

5. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;**31**: 166-9.

6. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;**11**: R106.

7. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;**5**: R80.

8. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods. 2015;**12**: 115-21.

9. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 2012;**40**: 4288-97.

10. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;**26**: 139-40.

11. Li J, Tibshirani R. Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. Stat Methods Med Res. 2013;**22**: 519-36.

12. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A. 2001;**98**: 5116-21.

13. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, et al. AmiGO: online access to ontology and annotation data. Bioinformatics. 2009;**25**: 288-9.

14. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res. 2013;**41**: D377-86.

15. Dietrich D, Uhl B, Sailer V, Holmes EE, Jung M, Meller S, et al. Improved PCR performance using template DNA from formalin-fixed and paraffin-embedded tissues by overcoming PCR inhibition. PLoS One. 2013;**8**: e77771.